July 14, 1997

Local Power of Likelihood Ratio Tests for the Cointegrating Rank of a VAR Process*

by

Pentti Saikkonen and Helmut Lütkepohl

Department of Statistics
University of Helsinki
P.O. Box 54
SF-00014 University of Helsinki
FINLAND

Tel.: +358-9-1918867
Fax: +358-9-1918872

Institut für Statistik und Ökonometrie
Wirtschaftswissenschaftliche Fakultät
Humboldt–University
Spandauer Str. 1
10178 Berlin
GERMANY
Tel.: +49-30-2093-5718
Fax: +49-30-2093-5712

Abstract

Likelihood ratio (LR) tests for the cointegrating rank of a vector autoregressive (VAR) process have been developed under different assumptions regarding deterministic terms. For instance, nonzero mean terms and linear trends have been accounted for in some of the tests. In this paper we provide a general framework for deriving the local power properties of these tests. Thereby it is possible to assess the virtue of utilizing varying amounts of prior information by making assumptions regarding the deterministic terms. One interesting result from this analysis is that if no assumptions regarding the specific form of the mean term are made while a linear trend is excluded then a test is available which has the same local power as an LR test derived under a zero mean assumption.

*We thank Ralf Brüggemann for performing the computations and the Deutsche Forschungsgemeinschaft, SFB 373, for financial support.

1 Introduction

Following the derivation of a full maximum likelihood (ML) analysis of cointegrated Gaussian vector autoregressive (VAR) processes by Johansen (1988, 1991a), likelihood ratio (LR) tests for the cointegrating rank have been developed under various sets of assumptions. The main differences in these assumptions relate to the deterministic terms such as intercept and mean terms as well as polynomial trends. In particular, LR tests for the cointegrating rank have been derived under the following conditions: (1) there is no deterministic term at all, (2) an intercept term is present only in the cointegration relations and there is no linear trend term, (3) a linear trend may be in the variables but not in the cointegration relations, (4) a linear trend is present in both the cointegration relations and in the variables, (5) an additive linear trend without any restrictions is added to the zero mean cointegrated stochastic part of the process. All these different assumptions result in different asymptotic null distributions of the LR tests. In this study we will derive the corresponding local power properties of the LR tests. These results enable us to assess the value of incorporating varying amounts of prior information included in the different sets of assumptions. Moreover, it is seen which factors are the crucial determinants of the local power of the tests. An important result is also that if an intercept term is present only in the cointegration relations and no linear trend is present in the process then a test can be constructed with identical local power to a test derived under scenario (1) where no deterministic term is present at all.

For some of the scenarios considered in this study, Johansen (1991b, 1995), Rahbek (1994) and Horvath & Watson (1995) have performed local power analyses. Our approach differs from that used in these articles, however. We will develop a general framework first in which the local power of the LR tests can be readily established.

This study is structured as follows. In the next section the model set-up is described and the LR tests are considered in Section 3. Since all these tests may be viewed as being obtained from a reduced rank (RR) regression a general result for such models is derived in Section 4. In Section 5 this result is used to obtain the local power of the LR tests for the cointegrating rank of a VAR process. Conclusions are given in Section 6 and proofs are contained in the Appendix.

The following notation is used throughout. The vector $y_t = (y_{1t}, \dots, y_{nt})'$ denotes an observable n-dimensional set of time series variables. The lag and differencing operators are denoted by L and Δ, respectively, that is, $Ly_t = y_{t-1}$ and $\Delta y_t = y_t - y_{t-1}$. The symbol $I(d)$ is used to denote a process which is integrated of order d, that is, it is stationary after differencing d times while it is still nonstationary after differencing just $d-1$ times. The symbols $\xrightarrow{d}$ and $\xrightarrow{p}$ signify convergence in distribution and probability, respectively, and a.s. is short for almost surely. $O(\cdot)$, $o(\cdot)$, $O_p(\cdot)$ and $o_p(\cdot)$ are the usual symbols for the order of convergence and convergence in probability, respectively, of a sequence. The normal distribution with mean (vector) μ and variance (covariance matrix) Σ is denoted by $N(\mu, \Sigma)$. The symbols $\lambda_{max}(A)$, $\mathrm{rk}(A)$ and $\mathrm{tr}(A)$ signify the maximal eigenvalue, the rank and the trace of the matrix A, respectively. If A is an $(n \times m)$ matrix of full column rank $(n > m)$ we let $A_\perp$ stand for an $(n \times (n-m))$ matrix of full column rank and such that $A' A_\perp = 0$. For an $(m \times n)$ matrix A and an $(m \times s)$ matrix B, $[A : B]$ is the $(m \times (n+s))$ matrix whose first n columns are the columns of A and whose last s columns are the columns of B. For a symmetric matrix A we write $A > 0$ to indicate that A is positive definite. The $(n \times n)$ identity matrix is denoted by I_n. LS is short for least squares and DGP abbreviates data generation process. RR means reduced rank. As a general convention, a sum is defined to be zero if the lower bound of the summation index exceeds the upper bound.

2 Preliminaries

Our point of departure is the DGP of an n-dimensional multiple time series $y_t = (y_{1t}, \dots, y_{nt})'$ defined by

$$y_t = \mu_0 + \mu_1 t + x_t, \qquad t = 1, 2 \dots, \tag{2.1}$$

where μ_0 and μ_1 are unknown, fixed $(n \times 1)$ parameter vectors and x_t is an unobservable error process with VAR(1) representation in error correction (EC) form

$$\Delta x_t = \Pi x_{t-1} + \varepsilon_t, \tag{2.2}$$

where $\varepsilon_t \sim iid\ N(0, \Omega)$, $x_0 = 0$ and Π is an $(n \times n)$ matrix of reduced rank r $(0 \leq r < n)$. Of course, this model set-up is simpler than in most applied studies with respect to the order of the process and the distribution of the residuals. The main reasons for choosing this simple model are that considering higher order short term dynamics makes the notation more complicated and has no impact on the results regarding the local power of those tests which are of primary interest in the following. It is also the framework used in other power studies to which we intend to compare our results (see Johansen (1995), Rahbek (1994)). The same is true for the assumption of normally distributed residuals. It is made mainly for convenience. Alternative distributional assumptions would have to be such that the same local power results are obtained and are therefore not of great interest for our purposes.

The rank of the matrix Π is the cointegrating rank of the variables x_t or, equivalently, of y_t. It is the focus of interest in the following. Suppose it is determined by testing the pair of hypotheses

$$H_0(r_0) : r = r_0 \qquad \text{vs.} \qquad H_1(r_0) : r > r_0. \tag{2.3}$$

It is also possible to consider the alternative hypothesis $H_1 : r = r_0 + 1$. For simplicity we will focus on $H_1(r_0)$ as given in (2.3) in this study. The local alternatives to be considered are given by

$$H_T(r_0) : \Pi = \alpha\beta' + T^{-1}\alpha_1\beta_1', \tag{2.4}$$

where α and β are fixed $(n \times r_0)$ matrices of rank r_0 and α_1 and β_1 are fixed $(n \times (r - r_0))$ matrices of rank $r - r_0$ and such that the matrices $[\alpha : \alpha_1]$ and $[\beta : \beta_1]$ have full column rank r. We use the assumption from Johansen (1995) and Rahbek (1994) that the eigenvalues of the matrix $I_{r_0} + \beta'\alpha$ are less than 1 in modulus.

Depending on the assumptions regarding the deterministic terms μ_0 and μ_1 there are different likelihood ratio tests for the hypotheses in (2.3). These tests will be reviewed in the next section.

3 Likelihood Ratio Tests

Most of the test statistics considered in this study may be obtained from reduced rank regressions of the from

$$\Delta y_t = \nu + \alpha B' y^*_{t-1} + z_t, \tag{3.1}$$

where ν is a fixed $(n \times 1)$ intercept vector, B is a suitable $(m \times r_0)$ matrix with $m \geq n$, y^*_{t-1} is an m-dimensional vector and z_t is an error term which contains all parts of the process which are not accounted for by the other quantities. The assumptions underlying the different tests amount to imposing restrictions on the intercept vector ν and choosing B and y^*_{t-1} appropriately. The following cases have been considered in the literature.

Case 1: $\mu_0 = \mu_1 = 0$, that is, the process has zero mean term and no linear trend. In this case the LR test statistic is obtained from a reduced rank regression

$$\Delta y_t = \alpha\beta' y_{t-1} + z_t,$$

that is, $\nu = 0$, $B = \beta$ and $y_{t-1}^* = y_{t-1}$ in (3.1). The resulting test statistic will be denoted by $LR^0(r_0)$. Critical values may be found in Johansen (1995, Table 15.1) or Reinsel & Ahn (1992, Table I) among others.

Case 2: μ_0 arbitrary, $\mu_1 = 0$, that is, there is no deterministic linear trend and this information is available. The test statistic is obtained from

$$\Delta y_t = \alpha(\beta' y_{t-1} + \delta) + z_t.$$

Hence, $\nu = 0$, $B' = [\beta' : \delta]$ and $y_{t-1}^* = [y_{t-1}' : 1]'$. The resulting test statistic will be denoted by $LR^*(r_0)$ and critical values may be found in Johansen (1995, Table 15.2). For this case Saikkonen & Luukkonen (1997) consider an alternative to the LR test which is based on constructing an estimator for μ_0 first, mean adjusting the data by subtracting that estimator and then applying an 'LR' test to the mean adjusted data. The resulting test statistics will be denoted by $LR^{SL}(r_0)$. It has the same limiting null distribution as $LR^0(r_0)$.

Case 3: μ_0 arbitrary, $\beta'\mu_1 = 0$, so that a linear trend may be present in the variables. In this case the relevant EC model for determining the test statistic is

$$\Delta y_t = \nu + \alpha\beta' y_{t-1} + z_t.$$

Thus, there is a nonzero intercept term, $B = \beta$ and $y_{t-1}^* = y_{t-1}$ in the framework of the general model (3.1). The asymptotic distribution of the LR statistic under $H_0(r_0)$ depends on whether or not $\mu_1 = 0$. Critical values for the case $\mu_1 = 0$ are given, e.g., in Johansen & Juselius (1990, Table A.2) or Reinsel & Ahn (1992, Table I). The test statistics used in conjunction with these critical values will be denoted by $LR^{i0}(r_0)$. Critical values for the situation where $\mu_1 \neq 0$ may be found, for example, in Johansen (1995, Table 15.3). The corresponding test statistics will be denoted by $LR^i(r_0)$.

Case 4: μ_0 and μ_1 arbitrary, that is, $\beta'\mu_1 \neq 0$ is possible. In that case a linear trend may be present in both the variables and the cointegrating relations. The relevant estimation equation is

$$\Delta y_t = \nu + \alpha(\beta' y_{t-1} + \tau(t-1)) + z_t.$$

Note that this model excludes quadratic trends without imposing restrictions on ν and τ. In the framework of (3.1) there is again a nonzero intercept term, $B' = [\beta' : \tau]$ and $y_{t-1}^* = [y_{t-1}' : t-1]'$. The test statistics will be denoted as $LR^+(r_0)$ and critical values may be obtained from Johansen (1995, Table 15.4).

Case 5: μ_0, μ_1 arbitrary and in estimating the trend parameters restrictions are not imposed to guarantee a linear trend. The difference to Case 4 is that the estimation is based on the equation

$$\Delta y_t = \nu_0 + \nu_1 t + \alpha\beta' y_{t-1} + z_t \tag{3.2}$$

which is not directly compatible with the model (3.1). It will be shown in the next section, however, that it can be treated in a similar way as the other cases. Without restrictions on ν_1 a model of the type (3.2) can generate quadratic deterministic trends. The resulting test was proposed by Perron & Campbell (1993) who derived the asymptotic properties of the test statistics which will be denoted by $LR^{PC}(r_0)$. Critical values may be found in Rahbek (1994) and Perron & Campbell (1993).

In the next section a general result will be given which allows to study the local power properties for the tests summarized here. The local power properties of $LR^0(r_0)$ are also given in Johansen (1991b, 1995) and those of $LR^{PC}(r_0)$ are derived in Rahbek (1994). Moreover, $LR^i(r_0)$ is known to have local power of a better order than the other tests (see again Rahbek (1994)). Thus, based on a local power criterion one would always apply $LR^i(r_0)$ if the underlying assumptions for this test can be justified. Unfortunately, in practice this may be difficult in many situations and one may consider using one of the other tests. Therefore we will compare the local power of those other tests in the following.

4 A General Result

We shall now give a general result for LR tests based on reduced rank (RR) regression. The following model will be considered:

$$Y_t = AB'X_t + Z_t, \qquad t = 1, \ldots, T, \tag{4.1}$$

where Y_t and Z_t are $(n \times 1)$ vectors, X_t is an $(m \times 1)$ vector with $m \geq n$ and A and B are $(n \times r_0)$ and $(m \times r_0)$ matrices of full column rank, respectively. The error term Z_t is of the form

$$Z_t = T^{-1} A_1 B_1' X_t + \mathcal{E}_t, \tag{4.2}$$

where A_1 and B_1 are $(n \times (r - r_0))$ and $(m \times (r - r_0))$ matrices, respectively, with $r - r_0 > 0$ and $\mathcal{E}_t$ is the error term under the null hypothesis that (4.1) is the correctly specified model. The matrices $[A : A_1]$ and $[B : B_1]$ are supposed to be of full column rank unless the null hypothesis holds, in which case $A_1 = 0$ and B_1 may also be zero. It may be worth noting that, in addition to the counterpart of the series Z_t, also the counterparts of the series Y_t, X_t and $\mathcal{E}_t$ may then depend on the sample size, as will be seen later. For ease of notation and because it has no effect on the general treatment in the following, we have not indicated the possible dependence of the quantities in (4.1) and (4.2) on the sample size.

As is well known, the RR estimators of A and B can be obtained as follows. First consider the generalized eigenvalues $\hat{\ell}_1 \geq \cdots \geq \hat{\ell}_n$ obtained as solutions of

$$\det(M_{XY} M_{YY}^{-1} M_{YX} - \ell M_{XX}) = 0, \tag{4.3}$$

where

$$M_{XX} = T^{-1} \sum_{t=1}^{T} X_t X_t', \quad M_{XY} = M_{YX}' = T^{-1} \sum_{t=1}^{T} X_t Y_t', \quad M_{YY} = T^{-1} \sum_{t=1}^{T} Y_t Y_t'.$$

Let $\hat{b}_1, \ldots, \hat{b}_n$ be the eigenvectors corresponding to $\hat{\ell}_1, \ldots, \hat{\ell}_n$ so that

$$(M_{XY} M_{YY}^{-1} M_{YX} - \hat{\ell}_j M_{XX}) \hat{b}_j = 0. \tag{4.4}$$

As usual, these eigenvectors are normalized as

$$\hat{b}_i' M_{XX} \hat{b}_j = \begin{cases} 1, & \text{if} \quad i = j \\ 0, & \text{if} \quad i \neq j \end{cases}. \tag{4.5}$$

Then we have $\hat{B} = [\hat{b}_1, \ldots, \hat{b}_r]$, while $\hat{A}$ is the LS estimator in a regression of Y_t on $\hat{B}' X_t$. Note that the foregoing formulation corresponds to that used by Johansen (1995, Section 6.1). Our first main result is the consistency of the RR estimators normalized in a suitable way. This result is obtained under the following general assumptions.

Assumption 1.

(i) $T^{-1} \sum_{t=1}^{T} B' X_t X_t' B \xrightarrow{p} \Sigma_{BB} > 0$

(ii) $T^{-1}\sum_{t=1}^{T} B_\perp' X_t X_t' B = O_p(1)$

(iii) $T^{-2}\sum_{t=1}^{T} X_t X_t' \xrightarrow{d} G$ for some (generally) random $(m \times m)$ matrix G with $B_\perp' G B_\perp > 0$ and $B'G = 0$ (a.s.)

(iv) $T^{-1/2}\sum_{t=1}^{T} \mathcal{E}_t X_t' B = O_p(1)$

(v) $T^{-1}\sum_{t=1}^{T} \mathcal{E}_t X_t' B_\perp \xrightarrow{d} S$ for some random $(n \times (m - r_0))$ matrix S

(vi) $T^{-1}\sum_{t=1}^{T} \mathcal{E}_t \mathcal{E}_t' = \Sigma_{\mathcal{E}\mathcal{E}} + O_p(T^{-1/2})$ for some fixed matrix $\Sigma_{\mathcal{E}\mathcal{E}} > 0$

Furthermore, the sequences in (iii) and (v) converge jointly in distribution.

The above formulation of the estimators enables us to mimic the consistency proof given in Johansen's (1995) Lemma 13.1. In the same way as in that lemma we also normalize the estimators $\hat{A}$ and $\hat{B}$ in a particular (infeasible) fashion to prove consistency. Consistency when other normalizations are used can then be obtained by the argument discussed in Johansen (1995, p. 180). Once the consistency of $\hat{A}$ and $\hat{B}$ has been proved it is easy to show that a consistent estimator of the matrix $\Sigma_{\mathcal{E}\mathcal{E}}$ is

$$\hat{\Sigma}_{\mathcal{E}\mathcal{E}} = T^{-1}\sum_{t=1}^{T}(Y_t - \hat{A}\hat{B}'X_t)(Y_t - \hat{A}\hat{B}'X_t)'. \tag{4.6}$$

The following lemma summarizes these results. It is shown in the Appendix.

Lemma 1

Consider the normalized estimators $\hat{B}_\gamma = \hat{B}(\gamma'\hat{B})^{-1}$ and $\hat{A}_\gamma = \hat{A}\hat{B}\gamma$, where $\gamma' = (B'B)^{-1}B'$. Then, if Assumption 1 holds, $\hat{B}_\gamma = B + O_p(T^{-1})$, $\hat{A}_\gamma = A + O_p(T^{-1/2})$ and $\hat{\Sigma}_{\mathcal{E}\mathcal{E}} = \Sigma_{\mathcal{E}\mathcal{E}} + O_p(T^{-1/2})$.

Let us now consider testing the null hypothesis that the RR regression equation (4.1) is correctly specified so that the error term Z_t equals $\mathcal{E}_t$. If $\mathcal{E}_t \sim iid\ N(0, \Sigma_{\mathcal{E}\mathcal{E}})$ and X_t is strictly exogenous or predetermined one can obtain the LR test against the alternative that the regression coefficient matrix is of full row rank. It can be shown that this test can be based on the auxiliary regression model

$$\hat{A}_\perp' Y_t = \Phi\hat{U}_t + R\hat{V}_t + N_t, \tag{4.7}$$

where $\hat{U}_t = \hat{B}'X_t$, $\hat{V}_t = \hat{B}'_\perp X_t$ and $N_t = \hat{A}'_\perp Z_t - \hat{A}'_\perp A(\hat{B} - B)'X_t$. Furthermore, $\Phi = \hat{A}'_\perp A$ and the true value of R is zero. The details are stated in the following lemma.

Lemma 2

The usual LR statistic for testing $H_0 : R = 0$ versus $H_1 : R \neq 0$ in the multivariate regression model (4.7) is identical to the LR statistic for testing $H_0 : \mathrm{rk}(\Psi) = r_0$ versus $H_1 : \mathrm{rk}(\Psi) > r_0$ in the Gaussian multivariate regression model $Y_t = \Psi X_t + \mathcal{E}_t$.

Of course, asymptotically equivalent tests can be obtained by using the corresponding Wald test or LM test. For convenience we will work with the LM version in the following. Hence, we consider the test statistic

$$LR(r_0) = \mathrm{tr}\{(\hat{A}'_\perp \hat{\Sigma}_{\mathcal{E}\mathcal{E}} \hat{A}_\perp)^{-1} \hat{R}\hat{M}_{V\cdot U}\hat{R}'\}, \qquad (4.8)$$

where $\hat{R}$ is the LS estimator of R from (4.7) and

$$\hat{M}_{V\cdot U} = \sum_{t=1}^{T} \hat{V}_t\hat{V}_t' - \sum_{t=1}^{T} \hat{V}_t\hat{U}_t' \left(\sum_{t=1}^{T} \hat{U}_t\hat{U}_t'\right)^{-1} \sum_{t=1}^{T} \hat{U}_t\hat{V}_t'. \qquad (4.9)$$

Notice that here we have assumed that the estimators used to construct the test statistic $LR(r_0)$ are obtained from the RR regression considered in Lemma 1. However, as far as asymptotic results are concerned, $\hat{A}$, $\hat{B}$ and $\hat{\Sigma}_{\mathcal{E}\mathcal{E}}$ can be any estimators for which the results of Lemma 1 hold. For instance, the Wald statistic is obtained by replacing $\hat{\Sigma}_{\mathcal{E}\mathcal{E}}$ in the definition of $LR(r_0)$ by

$$\tilde{\Sigma}_{\mathcal{E}\mathcal{E}} = T^{-1} \sum_{t=1}^{T} (Y_t - \tilde{\Psi}X_t)(Y_t - \tilde{\Psi}X_t)', \qquad (4.10)$$

where $\tilde{\Psi}$ is the full rank LS estimator of the product matrix AB' in (4.1). Now we are ready to state our main result.

Theorem 1

Suppose that Assumption 1 holds and $\hat{A}$, $\hat{B}$ and $\hat{\Sigma}_{\mathcal{E}\mathcal{E}}$ are any estimators satisfying the results of Lemma 1. Then, as $T \to \infty$,

$$\begin{aligned} LR(r_0) &\xrightarrow{d} \mathrm{tr}\{(A'_\perp \Sigma_{\mathcal{E}\mathcal{E}} A_\perp)^{-1}(A'_\perp A_1 B_1' G B_\perp + A'_\perp S)(B'_\perp G B_\perp)^{-1}(A'_\perp A_1 B_1' G B_\perp + A'_\perp S)'\} \\ &= \mathrm{tr}\{(A'_\perp \Sigma_{\mathcal{E}\mathcal{E}} A_\perp)^{-1}(F B'_\perp G B_\perp + A'_\perp S)(B'_\perp G B_\perp)^{-1}(F B'_\perp G B_\perp + A'_\perp S)'\}, \end{aligned}$$

where $F = A'_\perp A_1 B_1' B_\perp (B'_\perp B_\perp)^{-1}$.

Table 1. Relations of LR Test Statistics to RR Model (4.1).

Test statistic	Y_t	X_t	$\mathcal{E}_t$	A	B	A_1	B_1
$LR^0(r_0)$	Δy_t	y_{t-1}	ε_t	α	β	α_1	β_1
$LR^*(r_0)$	Δy_t	$[y'_{t-1} : 1]'$	ε_t	α	$[\beta' : \delta]'$	α_1	$[\beta'_1 : -\beta'_1\mu_0]'$
$LR^{SL}(r_0)$	$\Delta y_t - \overline{\Delta y}$	$y_{t-1} - \tilde{\mu}_0$	$\varepsilon_t + \alpha\beta'(\tilde{\mu}_0 - \mu_0)$ $+T^{-1}\alpha_1\beta'_1(\tilde{\mu}_0 - \mu_0)$	α	β	α_1	β_1
$LR^{i0}(r_0)$	$\Delta y_t - \overline{\Delta y}$	$y_{t-1} - \bar{y}_{-1}$	$\varepsilon_t - \bar{\varepsilon}$	α	β	α_1	β_1
$LR^+(r_0)$	$\Delta y_t - \overline{\Delta y}$	$\begin{bmatrix} y_{t-1} - \bar{y}_{-1} \\ t - 1 - \frac{1}{2}(T-1) \end{bmatrix}$	$\varepsilon_t - \bar{\varepsilon}$	α	$\begin{bmatrix} \beta \\ \tau \end{bmatrix}$	α_1	$\begin{bmatrix} \beta_1 \\ -\mu'_1\beta_1 \end{bmatrix}$
$LR^{PC}(r_0)$	$\Delta(y_t - \hat{\mu}_0 - \hat{\mu}_1 t)$	$y_{t-1} - \hat{\mu}_0 - \hat{\mu}_1(t-1)$	ε_t	α	β	α_1	β_1

Note: The overbar denotes the arithmetic mean. $\tilde{\mu}_0$ is an estimator of μ_0 which is described in Saikkonen & Luukkonen (1997). $\hat{\mu}_0$ and $\hat{\mu}_1$ are LS estimators of the trend parameters obtained from regressing y_t on 1 and t.

The proof of this result is also given in the appendix. Note that the limiting null distribution of the LR statistic is obtained by setting $A_1 = 0$. It may be worth noting that the limiting distribution depends on the random matrix S only through the term $A'_\perp S$. This fact will be useful later when explicit expressions of the asymptotic distribution in Theorem 1 are derived for special cases.

5 Local Power of LR Tests

5.1 Theory

The general result in Theorem 1 can be used to derive the asymptotic distributions of the LR statistics presented in Section 3 by writing the underlying model essentially in the form (4.1) and then showing that the relevant quantities Y_t, X_t and $\mathcal{E}_t$ satisfy the conditions summarized in Assumption 1. For the different test statistics the precise form of Y_t, X_t and $\mathcal{E}_t$ is given in Table 1. A specific form of each of the asymptotic distributions obtained from Theorem 1 is then derived for the individual tests using known limiting results. The following corollary gives the details. A full proof is given in the Appendix.

We use the following notation to state the results. The symbol $\mathbf{W}(u)$ is used to denote a Brownian motion with covariance matrix Ω and $\mathbf{K}(t)$ denotes the Ornstein-Uhlenbeck process defined by the integral equation

$$\mathbf{K}(u) = \alpha_\perp' \mathbf{W}(u) + \alpha_\perp' \alpha_1 \beta_1' \beta_\perp (\alpha_\perp' \beta_\perp)^{-1} \int_0^u \mathbf{K}(s) \mathrm{d}s \quad (0 \leq u \leq 1) \tag{5.1}$$

or, equivalently, the stochastic differential equation

$$\mathrm{d}\mathbf{K}(u) = \alpha_\perp' \mathrm{d}\mathbf{W}(u) + \alpha_\perp' \alpha_1 \beta_1' \beta_\perp (\alpha_\perp' \beta_\perp)^{-1} \mathbf{K}(u) \mathrm{d}u \quad (0 \leq u \leq 1)$$

(see, e.g., Johansen (1995, Chapter 14)). Furthermore, $\mathbf{N}(s)$ is the Ornstein-Uhlenbeck process defined by

$$\mathbf{N}(s) = (\alpha_\perp' \Omega \alpha_\perp)^{-1/2} \mathbf{K}(s) \quad \text{and} \quad \bar{\mathbf{N}}(s) = \mathbf{N}(s) - \int_0^1 \mathbf{N}(u) \mathrm{d}u. \tag{5.2}$$

Note that it is straightforward to check that alternatively $\mathbf{N}(s)$ may be defined as

$$\mathbf{N}(s) = \mathbf{B}(s) + ab' \int_0^s \mathbf{K}(u) \mathrm{d}u, \tag{5.3}$$

where $\mathbf{B}(s)$ is an $n - r$ dimensional standard Brownian motion and the quantities a and b are given by

$$a = (\alpha_\perp' \Omega \alpha_\perp)^{-1/2} \alpha_\perp' \alpha_1 \qquad \text{and} \qquad b = (\alpha_\perp' \Omega \alpha_\perp)^{1/2} (\beta_\perp' \alpha_\perp)^{-1} \beta_\perp' \beta_1 \tag{5.4}$$

[cf. Johansen (1995, pp. 207-208)]. In the following the argument of the Ornstein-Uhlenbeck processes is occasionally dropped when no confusion is possible. Now we can give the limiting distributions of the LR statistics under local alternatives.

Corollary 1

Under the assumptions for the DGP stated in Section 2 the following limiting results hold:

$$LR^0(r_0) \xrightarrow{d} \operatorname{tr}\left\{ \left(\int_0^1 \mathbf{N} \mathrm{d}\mathbf{N}' \right)' \left(\int_0^1 \mathbf{N}\mathbf{N}' \mathrm{d}s \right)^{-1} \left(\int_0^1 \mathbf{N} \mathrm{d}\mathbf{N}' \right) \right\},$$

$$LR^*(r_0) \xrightarrow{d} \operatorname{tr}\left\{ \left(\int_0^1 \mathbf{N}^* \mathrm{d}\mathbf{N}' \right)' \left(\int_0^1 \mathbf{N}^* \mathbf{N}^{*\prime} \mathrm{d}s \right)^{-1} \left(\int_0^1 \mathbf{N}^* \mathrm{d}\mathbf{N}' \right) \right\},$$

where $\mathbf{N}^*(s) = [\mathbf{N}(s)' : 1]'$,

$$LR^{SL}(r_0) \xrightarrow{d} \operatorname{tr}\left\{ \left(\int_0^1 \mathbf{N} \mathrm{d}\mathbf{N}' \right)' \left(\int_0^1 \mathbf{N}\mathbf{N}' \mathrm{d}s \right)^{-1} \left(\int_0^1 \mathbf{N} \mathrm{d}\mathbf{N}' \right) \right\},$$

$$LR^{i0}(r_0) \xrightarrow{d} \mathrm{tr}\left\{\left(\int_0^1 \bar{\mathbf{N}}\mathrm{d}\mathbf{N}'\right)'\left(\int_0^1 \bar{\mathbf{N}}\bar{\mathbf{N}}'\mathrm{d}s\right)^{-1}\left(\int_0^1 \bar{\mathbf{N}}\mathrm{d}\mathbf{N}'\right)\right\},$$

$$LR^{+}(r_0) \xrightarrow{d} \mathrm{tr}\left\{\left(\int_0^1 \mathbf{N}^{+}\mathrm{d}\mathbf{N}'\right)'\left(\int_0^1 \mathbf{N}^{+}\mathbf{N}^{+'}\mathrm{d}s\right)^{-1}\left(\int_0^1 \mathbf{N}^{+}\mathrm{d}\mathbf{N}'\right)\right\},$$

where $\mathbf{N}^{+}(s) = [\bar{\mathbf{N}}(s)' : s - \frac{1}{2}]'$, and

$$LR^{PC}(r_0) \xrightarrow{d} \mathrm{tr}\left\{\left(\int_0^1 \mathbf{N}^{PC}\mathrm{d}\mathbf{N}'\right)'\left(\int_0^1 \mathbf{N}^{PC}\mathbf{N}^{PC'}\mathrm{d}s\right)^{-1}\left(\int_0^1 \mathbf{N}^{PC}\mathrm{d}\mathbf{N}'\right)\right\},$$

where $\mathbf{N}^{PC}(s)$ is a trend adjusted version of $\mathbf{N}(s)$, that is, $\mathbf{N}(s)$ is corrected for mean and linear trend.

There are some interesting observations that can be made from this corollary. None of the limiting distributions depends on the dimension and cointegrating rank of the process directly but just on $n - r_0$, the number of common trends under the null hypothesis. Of course, this result is not surprising because it was also obtained for LR^0 and LR^{PC} by Johansen (1995) and Rahbek (1994). Moreover, it follows from (5.3) and (5.4) that the limiting distributions depend on α, β, Ω, α_1 and β_1 only through $a = (\alpha_\perp'\Omega\alpha_\perp)^{-1/2}\alpha_\perp'\alpha_1$ and $b = (\alpha_\perp'\Omega\alpha_\perp)^{1/2}(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'\beta_1$. This implies, for instance, for the case $r - r_0 = 1$, where α_1 and β_1 are $(n \times 1)$ vectors, that the limiting distributions depend on two parameters only, namely

$$f = b'a \quad \text{and} \quad g^2 = a'ab'b - (a'b)^2 \tag{5.5}$$

(see Johansen (1995, Corollary 14.5)). This fact is convenient in the simulations presented later.

The local power of the test statistics $LR^*(r_0)$, $LR^{SL}(r_0)$, $LR^{i0}(r_0)$, $LR^+(r_0)$ and $LR^{PC}(r_0)$, which allow for a nonzero mean μ_0, do not depend on the actual value of this mean term. Similarly the local power of none of the tests allowing for a linear trend depends on the actual value of the slope parameter vector μ_1.

Moreover, note that the limiting distribution of $LR^{SL}(r_0)$ is the same as that of $LR^0(r_0)$. This result was obtained by Saikkonen & Luukkonen (1997) under H_0 and is now seen to be valid also under local alternatives. It means that prior knowledge that $\mu_0 = 0$ is not helpful for improving the asymptotic local power of the test for the cointegrating rank. In other words, the same local power can be achieved with and without such prior knowledge.

For the univariate case, a similar result was also obtained by Elliott, Rothenberg & Stock (1996).

5.2 Simulations

Since the local power functions in Corollary 1 involve nonstandard distributions the relative efficiencies of the various tests are not obvious. Therefore, following Johansen (1995, Sec. 15.2), we have computed the local power for $r = r_0 + 1$ by simulating the discrete time counterpart of the Ornstein-Uhlenbeck process $\mathbf{N}(s)$. Note that from (5.3) we get $\mathrm{d}\mathbf{N}(u) = \mathrm{d}\mathbf{B}(u) + ab'\mathbf{N}(u)\mathrm{d}u$. Hence, in the simulations we use

$$\Delta \mathbf{N}_t = \frac{1}{T}\alpha_1\beta_1'\mathbf{N}_{t-1} + e_t, \qquad t = 1, \ldots, T = 1000,$$

with $e_t \sim iid\, N(0, I_{n-r_0})$, $\mathbf{N}_0 = 0$,

$$\beta_1' = \begin{cases} 1 & \text{for } n - r_0 = 1 \\ (1,0) & \text{for } n - r_0 = 2 \\ (1,0,0) & \text{for } n - r_0 = 3 \end{cases}$$

and

$$\alpha_1' = \begin{cases} f & \text{for } n - r_0 = 1 \\ (f,g) & \text{for } n - r_0 = 2 \\ (f,g,0) & \text{for } n - r_0 = 3 \end{cases}.$$

From these generated $\mathbf{N}_t$ we have computed

$$G_T = \frac{1}{T^2}\sum_{t=1}^{T} F_t F_t' \quad \text{and} \quad S_T = \frac{1}{T}\sum_{t=1}^{T} F_t \Delta \mathbf{N}_t',$$

where the definitions of the F_t for the different tests are given in Table 2. Finally, the values of the asymptotic LR statistics are obtained as $LR(r_0) = \mathrm{tr}(S_T' G_T^{-1} S_T)$. This experiment is repeated $R = 1000$ times and the resulting values of the test statistics are compared to the corresponding 5 % critical values of the relevant asymptotic null distributions. The relative rejection frequencies are depicted in Figures 1 - 4 for different values of f and g and different dimensions $n - r_0$.

A few interesting features can be seen in these figures. A first impression is that in general it pays to use as much prior information as possible. This result conforms with the conclusions from Horvath & Watson (1995) who analyze local power of LR tests in

Table 2. Definitions of F_t in Simulating Local Power

Test statistic	F_t
$LR^0(r_0)$	$\mathbf{N}_{t-1}$
$LR^*(r_0)$	$[\mathbf{N}_{t-1}' : 1]'$
$LR^{SL}(r_0)$	$\mathbf{N}_{t-1}$
$LR^{i0}(r_0)$	$\mathbf{N}_{t-1} - T^{-1}\sum_{t=1}^{T}\mathbf{N}_{t-1}$
$LR^+(r_0)$	$[(\mathbf{N}_{t-1} - T^{-1}\sum_{t=1}^{T}\mathbf{N}_{t-1})' : t-1-\frac{1}{2}(T-1)]'$
$LR^{PC}(r_0)$	$\mathbf{N}_{t-1} - \hat{\mu}_0 - \hat{\mu}_1(t-1)$

Note: $\hat{\mu}_0$ and $\hat{\mu}_1$ are LS estimators of the trend parameters obtained from regressing $\mathbf{N}_t$ on 1 and t.

the situation where some of the cointegrating vectors may be known. They also find that this kind of prior knowledge can result in substantial improvements in local power. Indeed, using knowledge regarding the deterministic terms can result in substantially more powerful tests in the present setting. For instance, $LR^{PC}(r_0)$ which assumes no knowledge regarding deterministic terms has much less power than $LR^0(r_0)$ which assumes knowledge that both μ_0 and μ_1 are zero. On the other hand, knowledge that the mean term is zero is not helpful for improving local power because $LR^{SL}(r_0)$ has the same local power as $LR^0(r_0)$ without using any knowledge on the mean term. It is striking, however, how much local power can be gained from estimating the mean term in the "right way" relative to just including an intercept term in the RR regression as in $LR^*(r_0)$ and $LR^{i0}(r_0)$. For many combinations of f and g the rejection probabilities of $LR^{SL}(r_0)$ are seen to be about twice as large as those of $LR^*(r_0)$ and $LR^{i0}(r_0)$. For instance, in Figure 1 for $f = -12$, the rejection frequency of $LR^{SL}(r_0)$ is 0.82 whereas $LR^*(r_0)$ and $LR^{i0}(r_0)$ have local power 0.31 and 0.45, respectively.

It is also interesting to see that, for a large part of the parameter space considered in our study, $LR^*(r_0)$ has smaller local power than $LR^{i0}(r_0)$, although both tests require the assumption that there is no deterministic trend term. This knowledge is used in $LR^*(r_0)$ to restrict the mean term to the cointegration relations whereas such a restriction is not used in $LR^{i0}(r_0)$. Obviously, in this case imposing the extra restriction in $LR^*(r_0)$ may result in a loss in asymptotic local power. This result is in line with the simulations of Horvath & Watson (1995) who compare the local power of $LR^{i0}(r_0)$ and $LR^*(r_0)$ in a more restrictive setting and find the same result. In fact, in Horvath & Watson's study $LR^*(r_0)$ was always inferior to $LR^{i0}(r_0)$. In Figures 2 and 3 it is seen that in part of our parameter space the opposite may be true. Of course, if $\mu_1 = 0$ is assumed so that there is no linear trend, then, from the point of view of local power maximization, neither $LR^*(r_0)$ nor $LR^{i0}(r_0)$ should be used. Clearly, $LR^{SL}(r_0)$ is the better choice in this case.

It is also interesting to compare the performance of $LR^+(r_0)$ and $LR^{PC}(r_0)$. The former test imposes the restriction that the estimated trend is at most linear whereas Perron & Campbell (1993) assume a linear trend in the DGP but do not impose this restriction in computing the test statistic $LR^{PC}(r_0)$. As a result the local power of the two tests differs. It can be seen in the figures, however, that $LR^+(r_0)$ is not always superior to $LR^{PC}(r_0)$ (see in particular Figure 1).

Another issue of practical importance is the dependence of the power on $n - r_0$, the number of stochastic trends under $H_0(r_0)$. In Figure 4 it is seen that increasing $n - r_0$ results in a loss of power for all the tests. This behaviour is not surprising. It was also observed by Johansen (1995) in studying the local power of $LR^0(r_0)$. He states that "the power decreases ... if there are many dimensions [for the additional cointegration vector] to hide in" (Johansen (1995, p. 213)).

5.3 Extensions

Notice that the test statistic $LR^*(r_0)$ can also be used for testing the joint hypothesis that $\Pi = \alpha\beta'$ and the intercept term $\nu = \alpha\delta$. In this set-up it may happen that the null hypothesis $\Pi = \alpha\beta'$ holds whereas $\nu \neq \alpha\delta$. In this case the intercept term in the model is unrestricted. This possibility was ruled out in Case 2 by assuming $\mu_1 = 0$. If $\nu = \alpha\delta$ were part of the null hypothesis it would be reasonable to consider also local alternatives of this part of the null hypothesis. Because these local alternatives would be of order $O(T^{-1/2})$ while those specified in $H_T(r_0)$ in (2.4) are of order $O(T^{-1})$, this case does not fit into our present framework. A similar comment applies with respect to the test statistic $LR^+(r_0)$.

6 Conclusions

We have investigated the asymptotic local power of LR tests for the cointegrating rank of a VAR process under various different assumptions regarding the properties of the deterministic terms. For this purpose a general framework for deriving the asymptotic distribution of LR tests under local alternatives has been presented. A number of LR tests for the cointegrating rank were then shown to fit into this framework and thus their local power properties could be established. The following main results have been obtained. (1) If the DGP is known to have no deterministic linear trend then the test suggested by Saikkonen & Luukkonen (1997) which is based on $LR^{SL}(r_0)$ is optimal from a local power point of view. It achieves the same power against local alternatives as the LR test which is based on the assumption that the DGP is known to have mean zero. (2) Not knowing whether there is possibly a linear trend and hence using $LR^{PC}(r_0)$ to be on the save side, results in a substantial loss of power in comparison with tests which are based on the assumption that no linear trend term is present. (3) The actual values of the trend and mean parameters do not enter the asymptotic distributions of the LR test statistics under local alternatives. Thus the actual magnitude of these parameters is of no relevance for the local power of these tests.

From a practical point of view it should perhaps be pointed out, however, that superior local power of a test does not necessarily imply superior power in small samples. Local power analysis is perhaps best thought of as an analysis of the power against alternatives close to the null hypothesis. Of course, achieving good power against such alternatives may be more important than good power against alternatives far away from the null for which it is relatively easy to determine that the null hypothesis is wrong anyway. In conclusion, while optimal local power is not a guarantee for optimal performance in all situations, tests with the former property are particularly useful in difficult situations where it is necessary to discriminate between nearby models. Hence, the local power properties should be a major factor in making a choice among different tests which may be available in a particular situation.

Appendix. Proofs

The notation from the previous sections of this paper is used here.

A.1 Proof of Lemma 1

First note that from (4.1), (4.2) and Assumption 1 one readily obtains

$$M_{YY} = AB'M_{XX}BA' + T^{-1}\sum_{t=1}^{T}\mathcal{E}_t\mathcal{E}_t' + o_p(1) = A\Sigma_{BB}A' + \Sigma_{\mathcal{E}\mathcal{E}} + o_p(1)$$

and

$$M_{YX}B = AB'M_{XX}B + o_p(1) = A\Sigma_{BB} + o_p(1).$$

Next, define $D_T = [B : T^{-1/2}B_\perp]$ and notice that (4.3) is equivalent to

$$\det(D_T'M_{XY}M_{YY}^{-1}M_{YX}D_T - \ell D_T'M_{XX}D_T) = 0. \tag{A.1}$$

This equation has the same eigenvalues as (4.3) and eigenvectors $D_T^{-1}\hat{b}_j$ $(j = 1, \ldots, n)$. As $T \to \infty$, the eigenvalues of $(A.1)$ converge weakly to those of the equation

$$\det(\Sigma_{BY}\Sigma_{YY}^{-1}\Sigma_{YB} - \ell\Sigma_{BB})\det(\ell B_\perp' G B_\perp) = 0,$$

where we have used the notation $\Sigma_{YY} = A\Sigma_{BB}A' + \Sigma_{\mathcal{E}\mathcal{E}}$ and $\Sigma_{YB} = \Sigma_{BY}' = A\Sigma_{BB}$. Thus, the situation is entirely analogous to that in the proof of Lemma 13.1 of Johansen (1995) and proceeding in the same way as there we can conclude that $\hat{B}_\gamma = B + o_p(T^{-1/2})$ and, furthermore, that $\hat{A}_\gamma = A + o_p(1)$ and $\hat{\Sigma}_{\mathcal{E}\mathcal{E}} = \Sigma_{\mathcal{E}\mathcal{E}} + o_p(1)$.

The next step is to establish the stated orders of consistency of $\hat{B}_\gamma$, $\hat{A}_\gamma$ and $\hat{\Sigma}_{\mathcal{E}\mathcal{E}}$. To this end, we write the first order conditions for $\hat{A}_\gamma$ and $\hat{B}_\gamma$ by modifying the analogs of Johansen's (1995) equations (13.8) and (13.9) in an obvious way after which the proof proceeds in the same way as in Johansen (1995, pp. 182-183) except that the relevant convergence results are obtained from Assumption 1 and the first part of the present proof. The last result of the lemma is not explicitly given by Johansen (1995) but it can be obtained in a straightforward manner from the order results for $\hat{A}_\gamma$ and $\hat{B}_\gamma$.

A.2 Proof of Lemma 2

Estimating the parameters of model (4.1) unrestrictedly by multivariate LS yields

$$Y_t = \tilde{\Psi}X_t + \tilde{\mathcal{E}}_t, \qquad t = 1, \ldots, T. \tag{A.2}$$

Let $\tilde{\Sigma}_{\mathcal{E}\mathcal{E}}$ be the corresponding estimator of the error covariance matrix $\Sigma_{\mathcal{E}\mathcal{E}}$ as in (4.10). Then the LR test statistic for $H_0 : \text{rk}(\Psi) = r_0$ can be written as

$$LR(r_0) = T \sum_{j=r_0+1}^{n} \log(1 + \hat{\lambda}_j), \tag{A.3}$$

where $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_n$ are the ordered generalized eigenvalues obtained as solutions of

$$\det(\tilde{\Psi} M_{XX} \tilde{\Psi}' - \lambda \tilde{\Sigma}_{\mathcal{E}\mathcal{E}}) = 0 \tag{A.4}$$

with M_{XX} as defined in Section 4. Let $\hat{\vartheta}_1 \geq \cdots \geq \hat{\vartheta}_n$ be the eigenvectors corresponding to $\hat{\lambda}_1, \ldots, \hat{\lambda}_n$ so that

$$(\tilde{\Psi} M_{XX} \tilde{\Psi}' - \hat{\lambda}_j \tilde{\Sigma}_{\mathcal{E}\mathcal{E}}) \hat{\vartheta}_j = 0. \tag{A.5}$$

These eigenvectors are normalized as

$$\hat{\vartheta}_i' \tilde{\Sigma}_{\mathcal{E}\mathcal{E}} \hat{\vartheta}_j = \begin{cases} 1, & \text{if} \quad i = j \\ 0, & \text{if} \quad i \neq j \end{cases} . \tag{A.6}$$

The Gaussian ML estimator of $B = [b_1, \ldots, b_r]$ is given by $\hat{B} = [\hat{b}_1, \ldots, \hat{b}_r]$, where

$$\hat{b}_j = \hat{\lambda}_j^{-1/2} \tilde{\Psi}' \hat{\vartheta}_j \qquad (j = 1, \ldots, n). \tag{A.7}$$

Note that $r = r_0$ under the null hypothesis. It follows from $(A.5) - (A.7)$ that we have the usual normalization $\hat{B}' M_{XX} \hat{B} = I_r$ as in Section 4 (see Anderson (1958, pp. 300 - 301)).

Let $\hat{\vartheta} = [\hat{\vartheta}_1, \ldots, \hat{\vartheta}_r]$, $\hat{\vartheta}_* = [\hat{\vartheta}_{r+1}, \ldots, \hat{\vartheta}_n]$ and $\hat{B}_* = [\hat{b}_{r+1}, \ldots, \hat{b}_n]$ with $r = r_0$ if the null hypothesis is assumed. Then multiplying $(A.2)$ by $[\hat{\vartheta} : \hat{\vartheta}_*]'$ gives

$$\hat{\vartheta}' Y_t = \hat{\Lambda}^{1/2} \hat{B}' X_t + \hat{\vartheta}' \tilde{\mathcal{E}}_t \tag{A.8}$$

$$\hat{\vartheta}_*' Y_t = \hat{\Lambda}_*^{1/2} \hat{B}_*' X_t + \hat{\vartheta}_*' \tilde{\mathcal{E}}_t \tag{A.9}$$

where $\hat{\Lambda} = \text{diag}[\hat{\lambda}_1, \ldots, \hat{\lambda}_r]$, $\hat{\Lambda}_* = \text{diag}[\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_n]$ and, by $(A.6)$, the residuals are uncorrelated (within the sample) with identity covariance matrix. The LR test statistic $(A.3)$ can clearly be obtained from $\hat{\Lambda}_*^{1/2}$ in $(A.9)$ without using the part of the model given in $(A.8)$. This shows that we may obtain the LR test statistic from a model which results from premultiplying $(A.2)$ by a suitable matrix.

To make this even more apparent, define $[\hat{\eta} : \hat{\eta}_*]' = [\hat{\vartheta} : \hat{\vartheta}_*]^{-1}$ and note that $(A.8)$ and $(A.9)$ imply

$$\begin{aligned} Y_t &= \hat{\eta} \hat{\Lambda}^{1/2} \hat{B}' X_t + \hat{\eta}_* \hat{\Lambda}_*^{1/2} \hat{B}_*' X_t + \tilde{\mathcal{E}}_t \\ &= \hat{A} \hat{B}' X_t + \hat{A}_* \hat{B}_*' X_t + \tilde{\mathcal{E}}_t, \end{aligned}$$

where $\hat{A} = \hat{\eta}\hat{\Lambda}^{1/2}$ and $\hat{A}_* = \hat{\eta}_*\hat{\Lambda}_*^{1/2}$. Thus, since $\hat{A}_\perp \in \text{span}(\hat{\vartheta}_*)$ it follows that the LR test statistic can also be obtained as follows. First the LS regression $(A.2)$ is premultiplied by the estimator $\hat{A}'_\perp$ and the regressor X_t is replaced by $\hat{B}'_* X_t$. Then the significance of the coefficient estimator of $\hat{B}'_* X_t$ is tested in the resulting (reduced) model (of $n - r_0$ equations) by using the conventional LR test of the multivariate linear model. Hence Lemma 2 is established.

A.3 Proof of Theorem 1

First note that the test statistic $LR(r_0)$ is invariant to normalizations of $\hat{A}, \hat{B}, \hat{A}_\perp$ and $\hat{B}_\perp$ so that we can assume that these estimators have been made unique by appropriate normalizations [cf. Lemma 1, Johansen (1995, Chapter 13) and Paruolo (1997)]. This implies also that the estimators $\hat{A}_\perp$ and $\hat{B}_\perp$ are consistent and their orders of consistency are $O_p(T^{-1/2})$ and $O_p(T^{-1})$, respectively. Next consider the estimator $\hat{R}$ and use standard LS theory to write

$$T\hat{R} = T^{-1}\sum_{t=1}^{T} N_t \hat{V}_t' (T^{-2}\hat{M}_{V\cdot U})^{-1}.$$

From Assumption 1 and the above mentioned consistency of $\hat{B}$ and $\hat{B}_\perp$ it readily follows that

$$T^{-2}\hat{M}_{V\cdot U} = T^{-2}\sum_{t=1}^{T} V_t V_t' + o_p(1) \xrightarrow{d} B'_\perp G B_\perp. \qquad (A.10)$$

Similar arguments and the definitions of N_t and Z_t give

$$\begin{aligned} T^{-1}\textstyle\sum_{t=1}^{T} N_t \hat{V}_t' &= T^{-1}\textstyle\sum_{t=1}^{T}(\hat{A}'_\perp Z_t - \hat{A}'_\perp A(\hat{B}-B)'X_t)\hat{V}_t' \\ &= T^{-1}\textstyle\sum_{t=1}^{T} \hat{A}'_\perp Z_t \hat{V}_t' + o_p(1) \\ &= \hat{A}'_\perp A_1 B_1' T^{-2}\textstyle\sum_{t=1}^{T} X_t \hat{V}_t' + \hat{A}'_\perp T^{-1}\textstyle\sum_{t=1}^{T} \mathcal{E}_t \hat{V}_t' + o_p(1) \\ &= A'_\perp A_1 B_1' T^{-2}\textstyle\sum_{t=1}^{T} X_t X_t' B_\perp + A'_\perp T^{-1}\textstyle\sum_{t=1}^{T} \mathcal{E}_t X_t' B_\perp + o_p(1). \end{aligned}$$

Thus, from Assumption 1(iii) and (v) we can conclude that

$$T^{-1}\sum_{t=1}^{T} N_t \hat{V}_t' \xrightarrow{d} A'_\perp A_1 B_1' G B_\perp + A'_\perp S. \qquad (A.11)$$

Furthermore, $(A.10)$ and $(A.11)$ in conjunction with the continuous mapping theorem yield

$$T\hat{R} \xrightarrow{d} (A'_\perp A_1 B_1' G B_\perp + A'_\perp S)(B'_\perp G B_\perp)^{-1}. \qquad (A.12)$$

Thus, since we obviously have $\hat{A}'_\perp \hat{\Sigma}_{\mathcal{E}\mathcal{E}} \hat{A}_\perp = A'_\perp \Sigma_{\mathcal{E}\mathcal{E}} A_\perp + o_p(1)$ the first form of the limiting distribution stated in Theorem 1 follows from $(A.11)$, $(A.12)$ and the continuous mapping theorem. The second form is obtained by using that $B_\perp(B'_\perp B_\perp)^{-1}B'_\perp + B(B'B)^{-1}B' = I_n$ and noting that $B'_1 G B_\perp = B'_1[B_\perp(B'_\perp B_\perp)^{-1}B'_\perp + B(B'B)^{-1}B']GB'_\perp = B'_1 B_\perp(B'_\perp B_\perp)^{-1}B'_\perp G B_\perp$ by Assumption 1(iii).

A.4 Proof of Corollary 1

As mentioned in Section 5, the limiting distributions given in Corollary 1 may be derived from Theorem 1 by writing the underlying model in the form (4.1) and then showing that Assumption 1 is satisfied. Finally, it is checked that the specific asymptotic distributions result from the general forms given in Theorem 1. Because the asymptotic distributions of $LR^0(r_0)$ and $LR^{PC}(r_0)$ have been derived by Johansen (1995, Chapter 14) and Rahbek (1994, Theorem 4.1) using a different approach we will not give detailed proofs of these results here to save space. Instead we begin by establishing the asymptotic distribution of $LR^{i0}(r_0)$.

A.4.1 Limiting Distribution of LR^{i0}

This test is obtained by a RR regression of the form

$$\overline{\Delta y_t} = \alpha\beta'\bar{y}_{t-1} + \bar{e}_t, \qquad t = 1, \ldots, T, \tag{A.13}$$

where the overbar signifies ordinary mean correction and the error term $\bar{e}_t$ has the representation

$$\bar{e}_t = T^{-1}\alpha_1\beta'_1\bar{y}_{t-1} + \bar{\varepsilon}_t. \tag{A.14}$$

Thus, we have a special case of (4.1) and (4.2) where the counterparts of Y_t, X_t and $\mathcal{E}_t$ ($\overline{\Delta y_t}$, $\bar{y}_{t-1}$ and $\bar{\varepsilon}_t$) depend on the sample size. To obtain the limiting distribution of the test statistic $LR^{i0}(r_0)$ from Theorem 1 it therefore suffices to check that Assumption 1 is satisfied.

Note that, since $\bar{y}_{t-1} = \bar{x}_{t-1}$ and $\Delta y_t = \Delta x_t$, we may assume that $\mu_0 = 0$. Let $\bar{y}_{-1}$ and $\bar{x}_{-1}$ be the sample means of y_{t-1} and x_{t-1}, respectively ($t = 1, \ldots, T$). By Theorem 14.1 of Johansen (1995) $\beta'\bar{y}_{-1}$ behaves asymptotically in the same way as under the null hypothesis $H_0(r_0)$. Thus, we have $\beta'\bar{y}_{-1} = O_p(T^{-1/2})$ and the validity of Assumption 1(i) follows from the second result of Johansen's (1995) Lemma 14.3. By Theorem 14.1 of Johansen (1995) and a standard application of the continuous mapping theorem we have

$$T^{-1/2}\alpha_\perp'\bar{y}_{-1} \xrightarrow{d} \int_0^1 \mathbf{K}(u)\mathrm{d}u$$

and, since we may assume that $\alpha_\perp = \beta c_1 + \beta_\perp$ (see below), it follows that here $\alpha_\perp$ can be replaced by $\beta_\perp$. Thus, we have $\beta_\perp'\bar{y}_{-1} = O_p(T^{1/2})$ and, by the argument given for Assumption 1(i), $\beta'\bar{y}_{-1} = O_p(T^{-1/2})$. These facts and the sixth result in Johansen's (1995) Lemma 14.3 imply that Assumption 1(ii) holds. Moreover, the following result is obtained from the fourth result of Johansen's (1995) Lemma 14.3:

$$T^{-2}\sum_{t=1}^{T}\beta_\perp'\bar{y}_{t-1}\bar{y}_{t-1}'\beta_\perp \xrightarrow{d} \int_0^1 \bar{\mathbf{K}}(s)\bar{\mathbf{K}}(s)'\mathrm{d}s, \tag{A.15}$$

where $\bar{\mathbf{K}}(s) = \mathbf{K}(s) - \int_0^1 \mathbf{K}(u)\mathrm{d}u$. From this result it readily follows that Assumption 1(iii) holds with $B_\perp' G B_\perp$ given by the right hand side of $(A.15)$.

As to Assumptions 1(iv) and (v), note first that they are known to hold under the null hypothesis $H_0(r_0)$. After this it is straightforward to conclude from Theorem 14.1 of Johansen (1995) and well-known properties of stationary and integrated processes that the same is true under $H_T(r_0)$. Next notice that

$$T^{-1}\sum_{t=1}^{T}\bar{\varepsilon}_t\bar{y}_{t-1}'\beta_\perp = T^{-1}\sum_{t=1}^{T}\varepsilon_t y_{t-1}'\alpha_\perp - T^{1/2}\bar{\varepsilon}T^{-1/2}\bar{y}_{-1}'\alpha_\perp + o_p(1).$$

The fifth result of Johansen's (1995) Lemma 14.3, the central limit theorem applied to $\bar{\varepsilon}$ and the limiting distribution of $T^{-1/2}\bar{y}_{-1}$ obtained above now show that

$$T^{-1}\sum_{t=1}^{T}\bar{\varepsilon}_t\bar{y}_{t-1}'\beta_\perp \xrightarrow{d} \int_0^1 \mathrm{d}\mathbf{W}(s)\bar{\mathbf{K}}(s)', \tag{A.16}$$

which means that Assumption 1(v) holds with S given by the right hand side of $(A.16)$. Finally, since the counterpart of $\mathcal{E}_t$ is $\bar{\varepsilon}_t$ the law of large numbers implies that Assumption 1(vi) holds with $\Sigma_{\mathcal{E}\mathcal{E}} = \Omega$.

Now, using these results the limiting distribution of the test statistic $LR^{i0}(r_0)$ under the local alternatives (2.4) can be deduced from Theorem 1, where the latter form of the limiting distribution is more convenient for our purposes. To be able to present the result in a convenient form we first note that the matrix $\beta_\perp'\beta_\perp$ can be replaced by $\beta_\perp'\alpha_\perp$. To see this, write $\alpha_\perp = \beta c_1 + \beta_\perp c_2$ and recall that the matrix $\beta_\perp'\alpha_\perp$ is nonsingular by assumption.

This implies that the matrix c_2 is nonsingular and, since the limiting distribution of the test statistic $LR^{i0}(r_0)$ is invariant to the transformation $\alpha_\perp \to \alpha_\perp c_2^{-1}$, we can assume that $c_2 = I_{n-r}$. Thus, we can write $\beta_\perp'\alpha_\perp = \beta_\perp'\beta_\perp$ so that the counterpart of the term $(FB_\perp'GB_\perp + A_\perp'S)'$ in Theorem 1 becomes

$$\int_0^1 \bar{\mathbf{K}}(s)\mathrm{d}\mathbf{W}(s)\alpha_\perp' + \int_0^1 \bar{\mathbf{K}}(s)\bar{\mathbf{K}}(s)'\mathrm{d}s(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'\beta_1\alpha_1'\alpha_\perp = \int_0^1 \bar{\mathbf{K}}(s)\mathrm{d}\mathbf{K}(s)', \qquad (A.17)$$

where the equality follows from the definition of the process $\mathbf{K}(s)$ [cf. Johansen (1995, p. 208)]. From this and our earlier discussion we can now conclude that $LR^{i0}(r_0)$ has the limiting distribution given in Corollary 1.

A.4.2 Limiting Distribution of LR^*

Now we turn to the proof of the limiting distribution of $LR^*(r_0)$. This test assumes that $\mu_1 = 0$ and it is based on the RR regression

$$\Delta y_t = \alpha\beta^{*'} y_{t-1}^* + e_t^*, \qquad t = 1, \ldots, T, \qquad (A.18)$$

where $y_t^* = [y_t' : 1]'$, $\beta^* = [\beta' : \delta]'$ with $\delta = -\beta'\mu_0$ and $e_t^* = \varepsilon_t + T^{-1}\alpha_1\beta_1^{*'}y_{t-1}^*$ with $\beta_1^* = [\beta_1' : \delta_1]'$ and $\delta_1 = -\beta_1'\mu_0$. For our purposes it will be convenient to reparameterize this model and consider instead the infeasible RR regression model

$$\Delta x_t = \alpha\beta_0^{*\prime} x_{t-1}^* + e_t^*, \qquad t = 1, \ldots, T, \qquad (A.19)$$

where $x_t^* = [x_t' : T^{1/2}]'$ and $\beta_0^{*\prime} = [\beta' : \delta_0]$ with $\delta_0 = T^{-1/2}(\beta'\mu_0 + \delta) = 0$. Of course, the error term e_t^* can be written accordingly as $e_t^* = \varepsilon_t + T^{-1}\alpha_1\beta_{10}^{*'}x_{t-1}^*$, where ${\beta_{10}^*}' = [\beta_1' : \delta_{10}]$ and $\delta_{10} = T^{-1/2}(\beta_1'\mu_0 + \delta_1) = 0$. Since $\Delta y_t = \Delta x_t$ it readily follows that the eigenvalues which appear in the test statistic $LR^*(r_0)$ can also be obtained from the infeasible model $(A.19)$. This model can therefore be used in theoretical considerations instead of $(A.18)$. Note that the square root of the sample size is used in x_t^*, δ_0 and δ_{10} to standardize the moment matrices in such a way that the RR regression in $(A.19)$ becomes conformable to what is required in Assumption 1. Clearly $(A.19)$ is a special case of (4.1) so that, to be able to apply Theorem 1, we have to verify Assumption 1. To this end, note that the counterparts of X_t, $\mathcal{E}_t$, B and B_1 are $x_{t-1}^*, \varepsilon_t, \beta_0^*$ and β_{10}^*, respectively. Since here β_0^* and β_{10}^* should be interpreted as "true" parameter values we have $\delta_0 = 0$ and $\delta_{10} = 0$ so that we may choose $\beta_{0\perp}^* = \mathrm{diag}[\beta_\perp : 1]$.

Since $\beta_0^{*\prime} x_{t-1}^* = \beta' x_{t-1}$, Assumption 1(i) holds by the second result of Johansen's (1995) Lemma 14.3, while Assumption 1(ii) follows from the sixth result of that lemma and the fact that the sample mean of the series $T^{1/2}\beta' x_{t-1}$ ($t = 1, \ldots, T$) is of order $O_p(1)$ by well-known properties of stationary processes. Next define $\xi_t^* = \beta_{0\perp}^{*\prime} x_{t-1}^* = [x_{t-1}'\beta_\perp : T^{1/2}]'$ and conclude from Theorem 14.1 of Johansen (1995) that $T^{-1/2}\xi_{[Ts]}^* \xrightarrow{d} \mathbf{K}^*(s)$, where $[Ts]$ denotes the integer part of Ts and $\mathbf{K}^*(s) = [\mathbf{K}(s)' : 1]'$. Note that here we can replace $\beta_\perp$ on the r.h.s. by $\alpha_\perp$ in the same way as in Section A.4.1. A standard application of the continuous mapping theorem now shows that

$$T^{-2}\sum_{t=1}^{T} \beta_{0\perp}^{*\prime} x_{t-1}^* x_{t-1}^{*\prime} \beta_{0\perp}^* \xrightarrow{d} \int_0^1 \mathbf{K}^*(s)\mathbf{K}^*(s)' \mathrm{d}s \tag{A.20}$$

by the fourth result of Johansen's (1995) Lemma 14.3 and by the result $T^{-1/2}\alpha_\perp' \bar{x}_{-1} \xrightarrow{d} \int_0^1 \mathbf{K}(u)\mathrm{d}u$ justified in Section A.4.1. The r.h.s. of (A.20) corresponds to $B_\perp' G B_\perp$ in Assumption 1(iii) and, thus, it follows that this part of Assumption 1 holds.

Assumption 1(iv) is again well-known under the null hypothesis and its validity under local alternatives can be obtained from Theorem 14.1 of Johansen (1995), (A.20) and well-known properties of stationary and integrated processes.

Furthermore,

$$\alpha_\perp' T^{-1}\sum_{t=1}^{T} \varepsilon_t x_{t-1}^{*\prime} \beta_{0\perp}^* \xrightarrow{d} \alpha_\perp' \int_0^1 \mathrm{d}\mathbf{W}(s)\mathbf{K}^*(s)', \tag{A.21}$$

where the r.h.s. is the counterpart of $\alpha_\perp' S$. To see this result and, hence, Assumption 1(v), just recall that $x_{t-1}^{*\prime}\beta_{0\perp}^* = [x_{t-1}'\beta_\perp : T^{1/2}]$, where $\beta_\perp$ can be replaced by $\alpha_\perp$, and apply the fifth result of Johansen's (1995) Lemma 14.3 augmented to include a constant. The validity of Assumption 1(vi) with $\Sigma_{\mathcal{E}\mathcal{E}}$ replaced by Ω is obvious because in place of $\mathcal{E}_t$ we have ε_t.

Using similar arguments as abvove, we can write $\beta_{0\perp}^{*\prime}\beta_{0\perp}^* = \mathrm{diag}[\beta_\perp'\alpha_\perp : 1]$ and, since now $\beta_{10}^* = [\beta_1' : 0]'$ we have $\beta_{0\perp}^{*\prime}\beta_{10}^* = [\beta_1'\beta_\perp : 0]'$. Thus, the counterpart of the term $(FB_\perp' G B_\perp + A_\perp' S)'$ in Theorem 1 becomes

$$\int_0^1 \mathbf{K}^*(s)\mathrm{d}\mathbf{W}(s)'\alpha_\perp + \int_0^1 \mathbf{K}^*(s)\mathbf{K}(s)'\mathrm{d}s(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'\beta_1\alpha_1'\alpha_\perp = \int_0^1 \mathbf{K}^*(s)\mathrm{d}\mathbf{K}(s)', \tag{A.22}$$

where the equality again follows from the definition of the process $\mathbf{K}(s)$. Hence, in the same way as in the case of test statistic $LR^{i0}(r_0)$ we can conclude that $LR^*(r_0)$ converges to the distribution given in the corollary.

A.4.3 Limiting Distribution of LR^+

Now consider the test statistic $LR^+(r_0)$ which is based on the RR regression model

$$\overline{\Delta y}_t = \alpha\beta^{+'} y_{t-1}^+ + e_t^+, \qquad t = 1, \dots, T, \tag{A.23}$$

where $y_t^+ = [\bar{y}'_{t-1} : \overline{(t-1)}]'$, $\beta^+ = [\beta' : \tau]'$ with $\tau = -\beta'\mu_1$ and $e_t^+ = \bar{\varepsilon}_t + T^{-1}\alpha_1\beta_1^{+'}y_{t-1}^+$ with $\beta_1^+ = [\beta_1' : \tau_1]'$ and $\tau_1 = -\beta_1'\mu_1$. It is again convenient to reparameterize $(A.23)$ and use instead the infeasible RR regression model

$$\overline{\Delta x}_t = \alpha\beta_0^{+'} x_{t-1}^+ + e_t^+, \qquad t = 1, \dots, T, \tag{A.24}$$

where $x_{t-1}^+ = [\bar{x}'_{t-1} : T^{-1/2}\overline{(t-1)}]'$, $\beta_0^+ = [\beta' : \tau_0]'$ with $\tau_0 = T^{1/2}(\beta'\mu_1 + \tau) = 0$ and $\overline{\Delta x}_t = \overline{\Delta y}_t$. The error term can correspondingly be rewritten as $e_t^+ = \bar{\varepsilon}_t + T^{-1}\alpha_1\beta_{10}^{+'}x_{t-1}^+$ with $\beta_{10}^+ = [\beta_1' : \tau_{10}]'$ and $\tau_{10} = T^{1/2}(\beta'\mu_1 + \tau_1) = 0$. In the same way as in the case of $(A.18)$ and $(A.19)$ the eigenvalues in $(A.23)$ and $(A.24)$ are identical so that the latter model, which is obviously a special case of (4.1), can be used to study theoretical properties of the test statistic $LR^+(r_0)$. Again we verify Assumption 1 and apply Theorem 1. For this purpose we define $\beta_0^+ = [\beta' : 0]'$ and $\beta_{10}^+ = [\beta_1' : 0]'$. Thus, we may again take $\beta_{0\perp}^+ = \text{diag}[\beta_\perp : 1]$.

Since $\beta_0^{+'}x_{t-1}^+ = \beta' x_{t-1} = \beta_0^{*'}x_{t-1}^*$ the validity of Assumption 1(i) and (ii) follow in the same way as in the case of test the statistic $LR^*(r_0)$. Furthermore, in the same way as in $(A.15)$ and $(A.20)$, we have

$$T^{-2}\sum_{t=1}^{T}\beta_{0\perp}^{+'}x_{t-1}^+x_{t-1}^{+'}\beta_{0\perp}^+ \xrightarrow{d} \int_0^1 \mathbf{K}^+(s)\mathbf{K}^+(s)'\mathrm{d}s = \int_0^1 \mathbf{K}^+(s)[\mathbf{K}(s)' : s]\mathrm{d}s, \tag{A.25}$$

where $\mathbf{K}^+(s) = [\bar{\mathbf{K}}(s)' : s - \frac{1}{2}]'$. The r.h.s. is the counterpart of the matrix $B_\perp' G B_\perp$ in Assumption 1(iii). This result follows by defining $\xi_t^+ = \beta_{0\perp}^{+'}x_{t-1}^+ = [\bar{x}'_{t-1}\beta_\perp : T^{-1/2}\overline{(t-1)}]'$. Then, using Theorem 14.1 of Johansen (1995), it is straightforward to show that $T^{-1/2}\xi_{[Tu]}^+ \xrightarrow{d} \mathbf{K}^+(u)$ and $(A.25)$ follows from a standard application of the continuous mapping theorem. As before we can thus conclude that Assumption 1(iii) holds. Assumption 1(iv) can again be justified by observing first that it is known to hold under the null hypothesis and then applying Johansen's (1995) Theorem 14.1 and well-known properties of stationary and integrated processes. Further notice that

$$T^{-1}\sum_{t=1}^{T}\bar{\varepsilon}_t x_{t-1}^{+'}\beta_{0\perp}^+ = T^{-1}\sum_{t=1}^{T}\varepsilon_t x_{t-1}^{+'}\beta_{0\perp}^+ = T^{-1}\sum_{t=1}^{T}\varepsilon_t \xi_{t-1}^{+\,\prime}.$$

Using the fifth result of Johansen's (1995) Lemma 14.3 and standard manipulations it can now be seen that

$$T^{-1}\sum_{t=1}^{T}\bar{\varepsilon}_t x_{t-1}^{+\prime}\beta_{0\perp}^{+} \xrightarrow{d} \int_0^1 \mathrm{d}\mathbf{W}(s)\mathbf{K}^{+}(s)' \tag{A.26}$$

and hence Assumption 1(v) holds. Finally Assumption 1(vi) is again obvious because in place of $\mathcal{E}_t$ we have $\bar{\varepsilon}_t$.

Since $\beta_{0\perp}^{+} = \beta_{0\perp}^{*}$ we can repeat the arguments below $(A.21)$ and conclude that the counterpart of the term $(FB_\perp' GB_\perp + A_\perp' S)'$ in Theorem 1 becomes

$$\int_0^1 \mathbf{K}^{+}\mathrm{d}\mathbf{W}(s)'\alpha_\perp + \int_0^1 \mathbf{K}^{+}(s)\mathbf{K}(s)'\mathrm{d}s(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'\beta_1\alpha_1'\alpha_\perp = \int_0^1 \mathbf{K}^{+}(s)\mathrm{d}\mathbf{K}(s)'. \tag{A.27}$$

Thus, in the same way as in the case of the test statistics $LR^{i0}(r_0)$ and $LR^{*}(r_0)$ we can conclude that $LR^{+}(r_0)$ has the limiting distribution stated in Corollary 1.

A.4.4 Limiting Distribution of LR^{SL}

Now consider the test statistic $LR^{SL}(r_0)$ which assumes that $\mu_1 = 0$ a priori and is based on the RR regression

$$\Delta\tilde{x}_t^{(0)} = \alpha\beta'\tilde{x}_{t-1}^{(0)} + \tilde{e}_t^{(0)}, \quad t = 1, \cdots, T, \tag{A.28}$$

where $\tilde{x}_t^{(0)} = y_t - \tilde{\mu}_0$, $\tilde{e}_t^{(0)} = \varepsilon_t + \alpha\beta'(\tilde{\mu}_0 - \mu_0) + T^{-1}\alpha_1\beta_1'(\tilde{\mu}_0 - \mu_0) + T^{-1}\alpha_1\beta_1'\tilde{x}_{t-1}^{(0)}$ and $\tilde{\mu}_0$ is a GLS estimator of the level parameter μ_0 described in Saikkonen & Luukkonen (1997). We will not give a detailed discussion of the estimator $\tilde{\mu}_0$ here but only concentrate on its main properties. The estimator $\tilde{\mu}_0$ is obtained in two steps of which the first one consists of computing the LS estimator of the parameter matrix Π in the EC model $\Delta y_t = \nu + \Pi y_{t-1} + \varepsilon_t$. This means running an LS regression of $\overline{\Delta y_t}$ on $\bar{y}_{t-1}$. The RR version of this LS regression was considered in Section A.4.1 (see Equation $(A.13)$) and Assumption 1 was verified for this case. Thus, it is straightforward to check that $\tilde{\Pi}$, the above mentioned LS estimator of Π, satisfies $(\tilde{\Pi} - \Pi)\beta = O_p(T^{-1/2})$ and $(\tilde{\Pi} - \Pi)\beta_\perp = O_p(T^{-1})$. These orders of consistency are exactly the same as under the null hypothesis so that following the arguments in the proof of Lemma 3.1 of Saikkonen & Luukkonen (1997) it can be shown that the GLS estimator $\tilde{\mu}_0$ has the properties $\beta'(\tilde{\mu}_0 - \mu_0) = O_p(T^{-1/2})$ and $\beta_\perp'(\tilde{\mu}_0 - \mu_0) = O_p(1)$. These results for the estimator $\tilde{\mu}_0$ are sufficient to obtain the limiting distribution of the test statistic $LR^{SL}(r_0)$ in the present context. Since $(A.28)$ is clearly a special case of (4.1) it suffices to verify Assumption 1 and apply Theorem 1. The counterparts of $X_t, \mathcal{E}_t, B$ and B_1 are obviously $\tilde{x}_{t-1}^{(0)}$, $\varepsilon_t + \alpha\beta'(\tilde{\mu}_0 - \mu_0) + T^{-1}\alpha_1\beta_1'(\tilde{\mu}_0 - \mu_0)$, β and β_1, respectively. Recall that $\tilde{x}_t^{(0)} = x_t - (\tilde{\mu}_0 - \mu_0)$ and that $\beta'(\tilde{\mu}_0 - \mu_0) = O_p(T^{-1/2})$ and $\beta_\perp'(\tilde{\mu}_0 - \mu_0) = O_p(1)$. Using these facts and results of the first and second sample moments of $\beta' x_{t-1}$ and $\beta_\perp' x_{t-1}$ already used in previous proofs it is straightforward to check that Assumption 1(i) and (ii) hold and also establish

$$T^{-2}\sum_{t=1}^{T}\beta_\perp'\tilde{x}_{t-1}^{(0)}\tilde{x}_{t-1}^{(0)\prime}\beta_\perp \xrightarrow{d} \int_0^1 \mathbf{K}(s)\mathbf{K}(s)'\mathrm{d}s \qquad (A.29)$$

which is the counterpart of the matrix $B_\perp' G B_\perp$ in Theorem 1. Thereby Assumption 1(iii) is shown to hold.

The verification of Assumption 1(iv) and (v) proceeds along similar lines. Therefore we consider the latter. We have to analyze

$$T^{-1}\sum_{t=1}^{T}\varepsilon_t\tilde{x}_{t-1}^{(0)\prime}\beta_\perp + \alpha\beta'(\tilde{\mu}_0 - \mu_0)T^{-1}\sum_{t=1}^{T}\tilde{x}_{t-1}^{(0)\prime}\beta_\perp + \alpha_1\beta_1'(\tilde{\mu}_0 - \mu_0)T^{-2}\sum_{t=1}^{T}\tilde{x}_{t-1}^{(0)\prime}\beta_\perp$$

$$\begin{aligned}
&= T^{-1}\sum_{t=1}^{T}\varepsilon_t x_{t-1}'\beta_\perp - T^{-1}\sum_{t=1}^{T}\varepsilon_t(\tilde{\mu}_0 - \mu_0)'\beta_\perp + \alpha\beta'(\tilde{\mu}_0 - \mu_0)T^{-1}\sum_{t=1}^{T}x_{t-1}'\beta_\perp \\
&- \alpha\beta'(\tilde{\mu}_0 - \mu_0)(\tilde{\mu}_0 - \mu_0)'\beta_\perp + \alpha_1\beta_1'(\tilde{\mu}_0 - \mu_0)T^{-2}\sum_{t=1}^{T}x_{t-1}'\beta_\perp \\
&- T^{-1}\alpha_1\beta_1'(\tilde{\mu}_0 - \mu_0)(\tilde{\mu}_0 - \mu_0)'\beta_\perp \\
&= T^{-1}\sum_{t=1}^{T}\varepsilon_t x_{t-1}'\beta_\perp + T^{1/2}\alpha\beta'(\tilde{\mu}_0 - \mu_0)T^{-3/2}\sum_{t=1}^{T}x_{t-1}'\beta_\perp + o_p(1).
\end{aligned}$$

The first term in the last expression converges weakly by the fifth result of Johansen's (1995) Lemma 14.3 while the (standardized) sum in the second term converges weakly by the argument given for the test statistic $LR^0(r_0)$. From the proof of Lemma 3.1 of Saikkonen & Luukkonen (1997) it can be seen that $T^{1/2}\beta'(\tilde{\mu}_0 - \mu_0)$ converges weakly and, since it is not difficult to check that all these weak convergencies hold jointly, we can conclude that Assumption 1(v) holds. Finally, since the counterpart of $\mathcal{E}_t$ is now $\varepsilon_t + \alpha\beta'(\tilde{\mu}_0 - \mu_0) + T^{-1}\alpha_1\beta_1'(\tilde{\mu}_0 - \mu_0)$ and $\alpha\beta'(\tilde{\mu}_0 - \mu_0) + T^{-1}\alpha_1\beta_1'(\tilde{\mu}_0 - \mu_0) = O_p(T^{-1/2})$ the validity of Assumption 1(vi) is immediate.

As for the counterpart of the matrix S, we will here concentrate on the transformed matrix $A_\perp' S$ which is obtained from

$$\alpha_\perp' T^{-1}\sum_{t=1}^{T}\varepsilon_t\tilde{x}_{t-1}^{(0)\prime}\beta_\perp \xrightarrow{d} \alpha_\perp'\int_0^1 \mathrm{d}\mathbf{W}(s)\mathbf{K}(s)'. \qquad (A.30)$$

In the same way as in $(A.17)$ the counterpart of the term $(FB_\perp' GB_\perp + A_\perp' S)'$ in Theorem 1 becomes

$$\int_0^1 \mathbf{K}(s)\mathrm{d}\mathbf{W}(s)'\alpha_\perp + \int_0^1 \mathbf{K}(s)\mathbf{K}(s)'\mathrm{d}s(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'\beta_1\alpha_1'\alpha_\perp = \int_0^1 \mathbf{K}(s)\mathrm{d}\mathbf{K}(s)'. \qquad (A.31)$$

Combining these results, the limiting distribution of the test statistic $LR^{SL}(r_0)$ is seen to be the same as that of the test statistic $LR^0(r_0)$. Hence, the corollary is established.

References

Anderson, T.W. (1958), *An Introduction to Multivariate Statistical Analysis*, John Wiley, New York.

Elliott, G., T.J. Rothenberg & J.H. Stock (1996), Efficient tests for an autoregressive unit root, *Econometrica*, 64, 813-836.

Horvath, M.T.K. & M.W. Watson (1995), Testing for cointegration when some of the cointegrating vectors are prespecified, *Econometric Theory*, 11, 984-1014.

Johansen, S. (1988), Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. (1991a), Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*, 59, 1551-1580.

Johansen, S. (1991b), The power function of the likelihood ratio test for cointegration, in J. Gruber (ed.), *Econometric Decision Models: New Methods of Modelling and Applications*, New York: Springer Verlag, pp. 323-335.

Johansen, S. (1995), *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*, Oxford: Oxford University Press.

Johansen, S. & K. Juselius (1990), Maximum likelihood estimation and inference on cointegration - with applications to the demand for money, *Oxford Bulletin of Economics and Statistics*, 52, 169-210.

Paruolo, P. (1997), Asymptotic inference on the moving average impact matrix in cointegrated I(1) VAR systems, *Econometric Theory*, 13, 79-118.

Perron, P. & J.Y. Campbell (1993), A note on Johansen's cointegration procedure when trends are present, *Empirical Economics*, 18, 777-789.

Rahbek, A.C. (1994), The power of some multivariate cointegration tests, Discussion Paper, Institute of Mathematical Statistics, University of Copenhagen.

Reinsel, G.C. & S.K. Ahn (1992), Vector autoregressive models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting, *Journal of Time Series Analysis*, 13, 353-375.

Saikkonen, P. & R. Luukkonen (1997), Testing cointegration in infinite order vector autoregressive processes, *Journal of Econometrics*, forthcoming.

Figure 1. Local power of LR tests for $n - r_0 = 1$.

Figure 2. Local power of LR tests for $n - r_0 = 2$.

Figure 3. Local power of LR tests for $n - r_0 = 3$.

Figure 4. Local power of LR tests for $g = 0$ and varying f and $n - r_0$.